

### REMARKS

Claims 1 through 12 are pending in this application.

Claims 1-12 were rejected.

Claim 1, 4, and 6 have been amended in this Response.

In the following, the Examiner's comments are included in bold, indented type, followed by the Applicant's remarks:

6. **Claims 1-3 and 6-12 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. Claims 1 and 10 recite "identifying R unique n-gram T1...R in the string; for every unique n-gram Ts: if the frequency of Ts in a set of n-gram statistics is not greater than a first threshold: associating the string with a cluster associated with Ts; otherwise: for every other n-gram Tv in the string T 1...R, except s: if the frequency of n-gram Tv is greater than the first threshold: if the frequency of n-gram pair Ts- Tv is not greater than a second threshold: associating the string with a cluster associated with the n-gram pair Ts-Tv; otherwise: for every other n-gram Tx in the string T1...R. except s and v: associating the string with a cluster associated with the n-gram triple Ts- Tv- Tx";**

**claim 6 recites "identifying R unique n-grams T1...R in the string; for every unique n-gram Ts: if the frequency of Ts in a set of n-gram statistics is not greater than a first threshold: associating the string with a cluster associated with Ts; otherwise: for i = 1 to Y: for every unique set of i n grams Tu in the string T1...R, except s: if the frequency of the n-gram set Ts-Tu is not greater than a second threshold: associating the string with a cluster associated with the n-gram set Ts Tu; if the string has not been associated with a cluster with this value of Ts: for every unique set of Y+1 n-grams Tuy in the string T1...R, except s: associating the string with a cluster associated with the Y+2 n-gram group Ts-Tuy". The specification page 6, line 19 through pages 8, line 15 as indicated by the Applicants does not provide any detail of the above-mentioned limitations of the claim.**

Applicants respectfully assert that the Examiner has not established that he has "a reasonable basis for questioning the adequacy of the disclosure to enable a person of ordinary skill in the art to make and use the claimed invention without resorting to *undue experimentation*." MPEP § 2106.01 (emphasis in original). In particular, the Examiner has not presented "a factual analysis of [the] disclosure to show that a person skilled in the art would not be able to make and use the claimed invention without resorting to undue experimentation." MPEP § 2106.02.

In any case, Applicants respectfully disagree. The claimed subject matter is generally described in the specification as originally filed on page 6, line 19 through page 8, line 15, which includes pseudocode (page 7, line 23 through page 8, line 12), and in Figs. 4, 5 and 6. An example is provided on page 8, lines 24-32 and Figs. 7, 8 and 9. An example SQL implementation flow for the string clustering technique is provided in Table 1 on page 9. Claims 1-12 are enabled because a person of ordinary skill in the art would know how to make and/or use the claimed invention based on these passages and the other material in the specification.

The pseudo-code on pages 7 and 8 uses different variable names from those used in the claims. Applicants do not know of any requirement and the Examiner has not cited any requirement that the specification use identical variable names as the claims. The figures generally omit the variable names used in the claim. Applicants are similarly unaware of any requirement that Applicants must use identical variable names in the figures and the claims. In any case, paragraphs [0004] – [0010] of the specification include variables with identical names to those used in the claims. The Examiner has not shown how the difference in variable naming between some portions of the specification and the claims renders the specification unable to “enable a person of ordinary skill in the art to make and use the claimed invention without resorting to *undue experimentation*.” MPEP § 2106.01 (emphasis in original).

8. Claims 4-5 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention. Claim 4, lines 5 and line 7 recites “if any”. Such language provides uncertainty or doubt, as whether the steps of associating each string with clusters will achieve. “if any” does not guarantee a completion of the associating step rather than a possibility of associating each string with clusters associated with low frequency n-grams from that string; and associating each string with clusters associated with low frequency pairs of high frequency n-grams from that string if it is existed. Applicant is advised to amend the claims to clarify that uncertainty set forth in the claims.

Applicants have amended claim 4 to clarify the claim. Applicants assert that claim 4, and dependent claim 5 are now definite and respectfully request that the Examiner remove the rejection of the claims.

10. Claims 1-3 and 6-9, are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

Claims 1-3 and 6-9, in view of MPEP section 2106 IV.B.2.(b) are not statutory because they merely recite a number of computing steps without producing any tangible result and/or being limited to a practical application within the technological arts. The language of the claim raises a question as to whether the claim is directed merely to an abstract idea that is not tied to a technological art, environment or machine which would result in a practical application producing a concrete, useful, and tangible result to form the basis of statutory subject matter under 35 U.S.C. 101.

With regarding claims 1,4 and 6:

While the preamble of the claim states, "a method for clustering a string including a plurality of characters", the claim fails to contain a computer that is used implemented the method for clustering a string so as to realize its functionality. Thus, claim 1 is merely abstract idea whereby "clustering a string including a plurality of characters" is being processed without any links to a practical result in the technology arts and without computer manipulation.

With regarding claims 2-3, 5 and 7-9:

The dependent claims 2-3, 5 and 7-9 are rejected for fully incorporating the errors of their respective base claims by dependency. Thus, claim 2-3, 5 and 7-9 are merely abstract idea and are being processed without any links to a practical result in the technology arts and without computer manipulation.

Applicants have amended claim 1, 4, and 6 to clarify that the claims are directed toward statutory subject matter. Applicants respectfully assert that claims 1, 4, and 6, as amended, and the claims that depend from claims 1, 4, and 6 comply with 35 U.S.C. § 101. Therefore, Applicants respectfully request that the Examiner remove the rejections of claims 1-9.

13. Claims 1-12 best understood by the examiner are rejected under 35 U.S.C. 103(a) as being unpatentable over Kreulen et al., (hereinafter "Kreulen") US Patent No. 6,862,586 and Chandrasekar et al., (hereinafter "Chandrasekar") US Patent no. 6,578,032. Claims 1, 6 and 10 can only be interpreted as best understood by the examiner. As to claims 1 and 10, Chandrasekar discloses "A method for clustering a plurality of strings, each string including a plurality of characters" as a use of providing a method for clustering character strings (col. 2, lines 23-25). In particular, Chandrasekar discloses the claimed "identifying R unique n-grams T1...R in the string" (col. 2, lines 3-10; col. 7, lines 30-45); "if the frequency of Ts in a set of n-gram statistics is not greater than a first threshold: associating the string with a cluster associated with Ts" (col. 12, line 59-col. 12, line 14); "for every other n-gram Tv in the string T1...R, except s: if the frequency of n-gram Tv is greater than the first threshold: if the frequency of n-gram pair Ts- Tv is not greater than a second threshold: associating the string with a cluster associated with the n gram pair Ts- Tv" (col. 12, line 59-col. 12, line 14). However, Chandrasekar does not explicitly disclose the use wherein "for every other n-gram Tx in the string T1...R. except s and v:

associating the string with a cluster associated with the n gram triple Ts-Tv- Tx". On the other hand, Kreulen discloses a method of searching a database using query, clustering the result items into logical categories and ranking the each categories based on the frequency of the occurrence of words (col. 1, line 67-col. 2, line 3). In particular, Kreulen discloses the claimed "for every other n-gram Tx in the string T1...R except s and v: associating the string with a cluster associated with the n-gram triple Ts-Tv- Tx" (col. 4, lines 50-56). Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the teachings of the cited references. One having ordinary skill in the art would have found it motivated to create an automated grouping using a clustering technique in order to provide easy update with the advent of a computer system.

Applicants respectfully disagree. The combination of Chandrasekar and Kreulen, assuming such a combination were possible, fails to teach each element of claim 1. Applicants will assume for this response that the Examiner's cited range of col. 12, line 59-col. 12, line 14 in Chandrasekar, should be col. 11, line 59-col 12, line 14. If this assumption is incorrect, Applicants respectfully request a corrected Office Action showing the correct range in Chandrasekar.

The portion of Chandrasekar in col. 11, line 59-col 12, line 14 discusses "one method of selecting a topic." Chandrasekar, col 11, lines 45-46. The portion of Chandrasekar in col. 11, line 59-col 12, line 14 generally discusses "calculating[ing] the frequency of the occurrence of the individual words and whole query" and determining "the highest frequency words and queries." Chandrasekar notes that "if none of the items satisfy a predetermined minimum threshold to become a topic, it may be that the longest item is selected to be the topic of the cluster." Chandrasekar, col. 12, lines 12-14. This is in contrast with "if the frequency of Ts in a set of n-gram statistics is not greater than a first threshold: associating the string with a cluster associated with Ts," as required by claim 1. The portion of Chandrasekar in col. 11, line 59-col 12, line 14 appears to teach away from this claim limitation by selecting a topic from the "highest frequency 'items' (i.e., words and/or queries)." Chandrasekar, col. 11, line 63.

Likewise, the portion of Chandrasekar in col. 11, line 59-col 12, line 14 fails to disclose or suggest the following limitations, as grouped by the Examiner:

for every other n-gram  $T_V$  in the string  $T_{1...R}$ , except S:

if the frequency of n-gram  $T_V$  is greater than the first threshold:

if the frequency of n-gram pair  $T_S-T_V$  is not greater than a second threshold:

associating the string with a cluster associated with the n-gram pair  $T_S-T_V$ .

The portion of Chandrasekar in col. 11, line 59-col 12, line 14 does not teach or suggest “the frequency of n-gram pairs.” A portion of Chandrasekar in col. 11, line 59-col 12, line 14 does say that “the two highest frequency items may be selected when their frequency scores are relatively close.” Chandrasekar, col 11, lines 65-66. Even assuming that the “two highest frequency items” are an n-gram pair within the meaning of the claim, the quoted section of Chandrasekar does not discuss the frequency of the pair as a pair, but rather discusses “the frequency of the occurrence of the individual words and whole query.” Chandrasekar, col. 11, lines 59-61, lines 65-66.

Applicants are unable to find any of the elements that include the decision-making element of “if the frequency of n-gram pair  $T_S-T_V$  is not greater than a second threshold” in the portion of Chandrasekar in col. 11, line 59-col 12, line 14. Therefore, the portion of Chandrasekar in col. 11, line 59-col 12, line 14 fails to disclose or suggest the claim limitations quoted above.

The cited portion of Kreulen discusses “represent[ing] [each document] as a triplet of unit vectors (D, F, B),” where “the first vector D in the triplet is the unit vector of normalized word frequencies for each word that occurs in the document (e.g., words, terms, or n-grams).” Kreulen, col. 4, lines 50-55. Kreulen later says that “the second vector F in the triplet is the unit vector of normalized out-link frequencies . . . [and] the third vector B in the triplet is the unit vector of normalized in-link frequencies.” Kreulen, col. 5, lines 1-2 and 11-12. Therefore, Kreulen's “triplet of unit vector (D, F, B)” is not an “n-gram triple  $T_S-T_V-T_X$ ” as referred to in

claim 1. Furthermore, the Examiner has not cited any portion of Kreulen, or any other reference, that shows how the creation of Kreulen's D vector is equivalent to "associating the string with a cluster." Therefore, the cited portion of Kreulen does not disclose or suggest "for every other n-gram  $T_X$  in the string  $T_{1...R}$ , except S and V: associating the string with a cluster associated with the n-gram triple  $T_S-T_V-T_X$ ."

Furthermore, the Examiner has not produced any reference that shows "if the frequency of n-gram  $T_v$  is greater than the first threshold . . . otherwise: do nothing," as required by claim 1.

As neither of the cited references, alone or in combination, disclose or suggest all of the limitations of claim 1, Applicants respectfully request that the Examiner withdraw his rejection of claim 1. Furthermore, because each of dependent claims 2 and 3 include all of the limitations of claim 1, which Applicants have shown to be patentable, Applicants respectfully request that the Examiner withdraw his rejection of claims 2 and 3.

As to claim 6, Chandrasekar discloses "A method for clustering a plurality of strings, each string including a plurality of characters" as a use of providing a method for clustering character strings (col. 2, lines 23-25). In particular, Chandrasekar discloses the claimed "identifying R unique n grams  $T_{1...R}$  in the string" (col. 2, lines 3-10; col. 7, lines 30-45); "if the frequency of  $T_s$  in a set of n-gram statistics is not greater than a first threshold" (col. 2, line 59-col. 2, line 14); "associating the string with a cluster associated with  $T_s$ ; otherwise: for  $i = 1$  to  $Y$ : for every unique set of  $i$  n-grams  $T_u$  in the string  $T_{1...R}$ , except  $s$ : if the frequency of the n-gram set  $T_s-T_u$  is not greater than a second threshold: associating the string with a cluster associated with the n gram set  $T_s-T_u$ " (col. 12, line 59-col. 12, line 14). However, Chandrasekar does not explicitly disclose the use wherein "if the string has not been associated with a cluster with this value of  $T_s$ : for every unique set of  $Y + 1$  n-grams  $T_{uy}$  in the string  $T_{1...R}$ , except  $s$ : associating the string with a cluster associated with the  $Y+2$  n-gram group  $T_s-T_{uy}$ ". On the other hand, Kreulen discloses a method of searching a database using query, clustering the result items into logical categories and ranking the each categories based on the frequency of the occurrence of words (col. 1, line 67-col. 2, line 3). In particular, Kreulen discloses the claimed "if the string has not been associated with a cluster with this value of  $T_s$ : for every unique set of  $Y + 1$  n-grams  $T_{uy}$  in the string  $T_{1...R}$ , except  $s$ : associating the string with a cluster associated with the  $Y+2$  n-gram group  $T_s-T_{uy}$ " (col. 4, lines 50-56). Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the teachings of the cited references. One having ordinary skill in the art would have found it motivated to create an automated grouping

**using a clustering technique in order to provide easy update with the advent of a computer system.**

Applicants respectfully disagree. The combination of Chandrasekar and Kreulen, assuming such a combination were possible, fails to teach each element of claim 6. Applicants will assume for this response that the Examiner's cited range of col. 12, line 59-col. 12, line 14 in Chandrasekar, should be col. 11, line 59-col 12, line 14. If this assumption is incorrect, Applicants respectfully request a corrected Office Action showing the correct range in Chandrasekar.

The portion of Chandrasekar in col. 11, line 59-col 12, line 14 discusses "one method of selecting a topic." Chandrasekar, col 11, lines 45-46. The portion of Chandrasekar in col. 11, line 59-col 12, line 14 generally discusses "calculating[ing] the frequency of the occurrence of the individual words and whole query" and determining "the highest frequency words and queries." Chandrasekar notes that "if none of the items satisfy a predetermined **minimum threshold** to become a topic, it may be that the longest item is selected to be the topic of the cluster." Chandrasekar, col. 12, lines 12-14. This is in contrast with "if the frequency of  $T_s$  in a set of n-gram statistics **is not greater than** a first threshold: associating the string with a cluster associated with  $T_s$ ," as required by claim 6. The portion of Chandrasekar in col. 11, line 59-col 12, line 14 appears to teach away from this claim limitation by selecting a topic from the "highest frequency 'items' (i.e., words and/or queries)." Chandrasekar, col. 11, line 63.

Likewise, the portion of Chandrasekar in col. 11, line 59-col 12, line 14 fails to disclose or suggest the following limitations, as grouped by the Examiner:

associating the string with a cluster associated with  $T_S$ ;  
otherwise:  
for  $i = 1$  to  $Y$ :  
for every unique set of  $i$  n-grams  $T_U$  in the string  $T_{1...R}$ , except  $S$ :  
if the frequency of the n-gram set  $T_S - T_U$  is not greater than a second threshold:  
associating the string with a cluster associated with the n-gram set  $T_S - T_U$ ;

The portion of Chandrasekar in col. 11, line 59-col 12, line 14 does not discuss “the frequency of the n-gram set  $T_S - T_U$ .” A portion of Chandrasekar in col. 11, line 59-col 12, line 14 says that “the two highest frequency items may be selected when their frequency scores are relatively close.” Chandrasekar, col 11, lines 65-66. Even assuming that the “two highest frequency items” are an n-gram set within the meaning of the claim, the quoted section of Chandrasekar does not discuss the frequency of the set as a set, but rather discusses “the frequency of the occurrence of the individual words and whole query.” Chandrasekar, col. 11, lines 59-61, lines 65-66. Therefore, the portion of Chandrasekar in col. 11, line 59-col 12, line 14 does not disclose or suggest the claim limitations quoted above. Furthermore, Applicants are unable to find any disclosure or suggestion of “associating the string with a cluster associated with the n-gram set  $T_S - T_U$ ,” as required by claim 6 in the portion of Chandrasekar in col. 11, line 59-col 12, line 14.

The cited portion of Kreulen discusses “represent[ing] [each document] as a triplet of unit vectors (D, F, B),” where “the first vector D in the triplet is the unit vector of normalized word frequencies for each word that occurs in the document (e.g., words, terms, or n-grams).” Kreulen, col. 4, lines 50-55. Kreulen later says that “the second vector F in the triplet is the unit vector of normalized out-link frequencies . . . [and] the third vector B in the triplet is the unit vector of normalized in-link frequencies.” Kreulen, col. 5, lines 1-2 and 11-12. Therefore, Kreulens “triplet of unit vector (D, F, B)” is not a “Y+2 n-gram group  $T_S - T_{UY}$ ” as required by



claim 6. Furthermore, the Examiner has not cited any portion of Kreulen, or any other reference, that shows how the creation of Kreulen's D vector or any other cited portion of Kreulen is equivalent to "associating the string with a cluster." Therefore, the cited portion of Kreulen does not disclose or suggest "if the string has not been associated with a cluster with this value of  $T_S$ : for every unique set of  $Y+1$  n-grams  $T_{UY}$  in the string  $T_{1...R}$ , except  $S$ : associating the string with a cluster associated with the  $Y+2$  n-gram group  $T_S-T_{UY}$ ."

As neither of the cited references, alone or in combination, disclose or suggest all of the limitations of claim 6, Applicants respectfully request that the Examiner withdraw his rejection of claim 6. Furthermore, because each of dependent claims 7-9 include all of the limitations of claim 6, which Applicants have shown to be patentable, Applicants respectfully request that the Examiner withdraw his rejection of claims 2 and 3.

As to claim 4, Chandrasekar discloses "A method for clustering a plurality of strings, each string including a plurality of characters" as a use of providing a method for clustering character strings (col. 2, lines 23-25). In particular, Chandrasekar discloses the claimed "identifying unique n grams in each string" (col. 2, lines 3-10; col. 1, lines 30-45). However, Chandrasekar does not explicitly disclose the use of associating each string with clusters associated with low frequency n-grams from that string, if any and associating each string with clusters associated with low frequency pairs of high frequency n-grams from that string, if any. On the other hand, Kreulen discloses a method of searching a database using query, clustering the result items into logical categories and ranking the each categories based on the frequency of the occurrence of words (col. 1, line 67-col. 2, line 3). In particular, Kreulen discloses the claimed "associating each string with clusters associated with low frequency n-grams from that string, if any" (col. 4, lines 50-56); and "associating each string with clusters associated with low-frequency pairs of high frequency n-grams from that string, if any" (col. 4, lines 50-56). Therefore, it would have been obvious to one having ordinary skill in the art at the time the invention was made to combine the teachings of the cited references, wherein the Editorial database, provided therein (see Chandrasekar's fig. 8) would incorporate the use of associating each string with clusters associated with low frequency n-grams from that string, if any and associating each string with clusters associated with low-frequency pairs of high frequency n-grams from that string, if any, in the same conventional manner as disclosed by Kreulen(col. 4, lines 50-56). One having ordinary skill in the art would have found it motivated to create an automated grouping using a clustering technique in order to provide easy update with the advent of a computer system.

Applicants respectfully disagree. The combination of Chandrasekar and Kreulen, assuming such a combination were possible, fails to teach each element of claim 4. The cited portion of Kreulen discusses “represent[ing] [each document] as a triplet of unit vectors (D, F, B),” where “the first vector D in the triplet is the unit vector of normalized word frequencies for each word that occurs in the document (*e.g.*, words, terms, or n-grams).” Kreulen, col. 4, lines 50-55. Kreulen later says that “the second vector F in the triplet is the unit vector of normalized out-link frequencies . . . [and] the third vector B in the triplet is the unit vector of normalized in-link frequencies.” Kreulen, col. 5, lines 1-2 and 11-12. Therefore, Kreulen’s “triplet of unit vector (D, F, B)” is not “zero or more low-frequency pairs of high frequency n-grams from that string,” as required by claim 4. Furthermore, the Examiner has not cited any portion of Kreulen, or any other reference, that shows how the creation of Kreulen’s D vector or any other cited portion of Kreulen is equivalent to “associating each string with zero or more clusters associated with low frequency n-grams from that string; and associating each string with clusters associated with zero or more low-frequency pairs of high frequency n-grams from that string,” as required by claim 4. Therefore, the cited portion of Kreulen does not disclose or suggest each of the elements of claim 4.

As neither of the cited references, alone or in combination, disclose or suggest all of the limitations of claim 4, Applicants respectfully request that the Examiner withdraw the rejection of claim 4. Furthermore, because dependent claims 5 includes all of the limitations of claim 4, which Applicants have shown to be patentable, Applicants respectfully request that the Examiner withdraw the rejection of claims 5.

**SUMMARY**

Applicants contend that the claims are in condition for allowance, which action is requested. Applicants do not believe any fees are necessary with the submitting of this response. Should any fees be required, Applicants request that the fees be debited from deposit account number 50-1673.

Respectfully submitted,



Howard L. Speight

Reg. No. 37,733

Baker Botts L.L.P.

910 Louisiana

Houston, Texas 77002

Telephone: (713) 229-2057

Facsimile: (713) 229-2757

E.Mail: Howard.Speight@bakerbotts.com

ATTORNEY FOR APPLICANT

Date: July 28, 2005

## AMENDMENTS TO THE DRAWINGS

Regarding the drawings, the Examiner stated:

**The drawings are objected to under 37 CFR 1.83(a). The drawings must show every feature of the invention specified in the claims. Therefore, the “identifying R unique n-gram T1...R in the string; for every unique n-gram Ts: if the frequency of Ts in a set of n-gram statistics is not greater than a first threshold: associating the string with a cluster associated with Ts; otherwise: for every other n-gram Tv in the string T1...R, except s: if the frequency of n gram pair Tv is greater than the first threshold: if the frequency of n-gram pair Ts- Tv is not greater than a second threshold: associating the string with a cluster associated with the n gram pair Ts-Tv; otherwise: for every other n-gram Tx in the string T1...R, except s and v: associating the string with a cluster associated with the n-gram triple Ts-Tv-Tx” and “identifying R unique n-grams T1...R in the string; for every unique n-gram Ts: if the frequency of Ts in a set of n-gram statistics is not greater than a first threshold: associating the string with a cluster associated with Ts; otherwise: for i = 1 to Y: for every unique set of i n-grams Tu in the string T1...R, except s: if the frequency of the n-gram set Ts-Tu is not greater than a second threshold: associating the string with a cluster associated with the n-gram set Ts- Tu; if the string has not been associated with a cluster with this value of Ts: for every unique set of Y + 1 n grams Tuy in the string T1...R, except s: associating the string with a cluster associated with the Y+ 2 n gram group Ts-Tuy” must be shown or the feature(s) canceled from the claim(s). No new matter should be entered.**

Applicants are submitting additional figures numbered 10A, 10B, 11A, 11B, and 11C with this response. Each figure is supported by disclosure in the application as originally filed. Specifically, Figures 10A and 10B are disclosed in claim 1 as originally filed, Figures 4, 5, 6A, and 6B, paragraphs [0004], [0025]-[0034] of the specification, pseudo-code on page 7, line 23-page 8, line 12, and example SQL statements in Table 1 on page 9. Specifically, Figures 11A, 11B, and 11C are disclosed in claim 6 as originally filed, Figures 4, 5, 6A, and 6B, paragraphs [0008]-[0009], [0025]-[0034] of the specification, pseudo-code on page 7, line 23-page 8, line 12, and example SQL statements in Table 1 on page 9.